

1. Analisi esplorative

1.1 Analisi Univariata

Il database è composto da 44 osservazioni, una per ciascun fiume, e 9 variabili. Le variabili rappresentano varie caratteristiche dei fiumi come il numero di specie di cozze presenti, la dimensione del fiume, la distanza da alcuni fiumi importanti, e vari indicatori sulla qualità delle acque.

	Min.	1 ^o qt.	Mediana(IQR)	Media(sd)	3 ^o qt.	Max
Species	2.00	8.00	10.00(5.0)	11.25(5.99)	13.00	33.00
Area	349	2115	4315(7797.5)	6590(6016)	9912	27900
Stepping						
stones to AC	1.00	7.00	15.50(15.25)	15.34(9.19)	22.25	33.00
Stepping						
stones to AP	0.00	4.00	12.00(14.25)	12.02(8.24)	18.25	28.00
Stepping						
stones to SV	0.00	5.00	7.00(13.25)	8.136(8.94)	11.00	21.00
Stepping						
stones to SL	4.00	16.75	22.00(0.775)	22.16(1.84)	30.00	36.00
Nitrate	0.100	0.600	0.800(0.775)	1.495(1.84)	1.375	8.700
Solid Residue	29.0	56.5	78.0(64.0)	112.4(17.3)	120.5	520.0
Hydronium	0.200	1.00	1.600(2.2)	3.631(6.03)	3.200	32.00
ln(Area)	5.855	7.657	8.370(1.54)	8.331(1.07)	9.201	10.24

Tabella 1: Tabella di sintesi

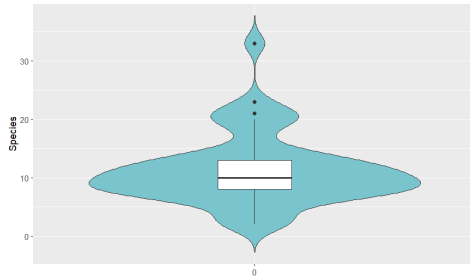


Figura 1: Violinplot della variabile
species

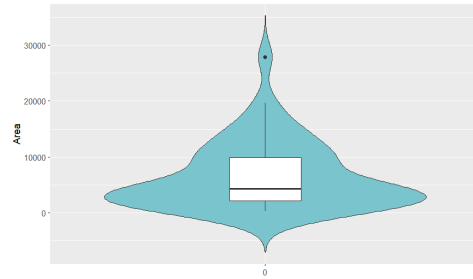


Figura 2: Violinplot della variabile
Area

La variabile **Species**, rappresentante il numero di specie di cozze nel fiume, ha una distribuzione asimmetrica a destra come evidenziato dal violinplot (*Figura 1*). Ha un valore minimo di 2 specie (nel fiume Waccasassa) e un massimo di 33 specie (nel fiume Apalachicola). La mediana è di 10 specie, la media è di 11.25 specie. La variabile **Area** segue una distribuzione asimmetrica a destra (*Figura 2*), con un valore minimo di 349 sq mi (*square mile*) e un massimo di 27900 sq mi. La mediana è di 4315 sq mi, la media è di 6590 sq mi (*Tabella 1*).

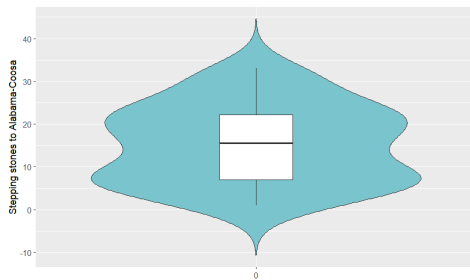


Figura 3: Violinplot della variabile
stepping stones to AC

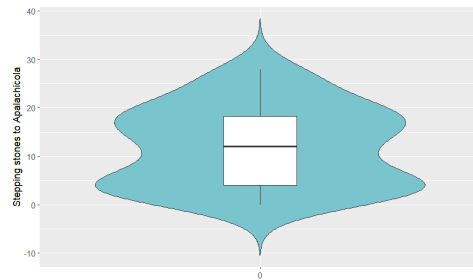


Figura 4: Violinplot della variabile
stepping stones to AP

La distanza dal sistema fluviale di Alabama-Coosa (**Stepping stones to AC**) segue una distribuzione simmetrica (*Figura 3*), suggerendo che i fiumi sono equamente distribuiti rispetto a questo sistema fluviale. La distanza minima è di 1 e la massima di 33. La distanza media è di 15.34, e quella mediana di 15.50 "saltelli". Anche la distanza dal fiume Apalachicola (**Stepping stones to AP**) segue una distribuzione simmetrica (*Figura 4*), con una distanza minima di 0 (coincidente con il fiume stesso) e una massima di 28. La distanza media è di 12.02, e quella mediana di 12.00 "saltelli" (*Tabella 1*).

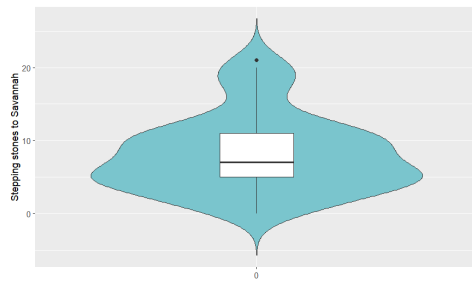


Figura 5: Violinplot della variabile
stepping stones to SV

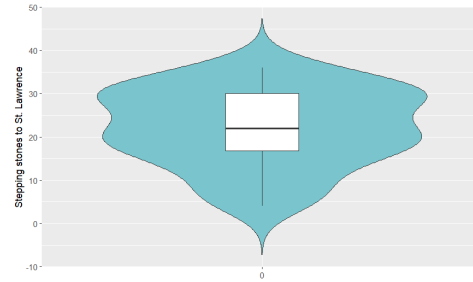


Figura 6: Violinplot della variabile
stepping stones to SL

La distanza dal fiume Savannah (**Stepping stones to SV**) ha una distribuzione simmetrica anche se presenta una leggera coda lunga a destra (*Figura 5*), con una distanza minima di 0 (coincidente con il fiume stesso) e una massima di 21. La distanza media è di 8.136, e quella mediana di 7.00. Il valore molto basso della mediana e la presenza di una piccola coda lunga a destra suggeriscono che la maggior parte dei fiumi osservati sono molto vicini al fiume Savannah. La distanza dal fiume St. Lawrence (**Stepping stones to SL**) ha una distribuzione simmetrica (*Figura 6*), ma presenta anche una leggera coda lunga a sinistra. La distanza minima di 4 e una massima di 36. La distanza media è di 22.16, e quella mediana di 22.00. Il valore elevato della di media e mediana, in aggiunta alla leggera coda a sinistra, suggerisce che la maggior parte dei fiumi sono distanti al fiume St. Lawrence (*Tabella 1*).

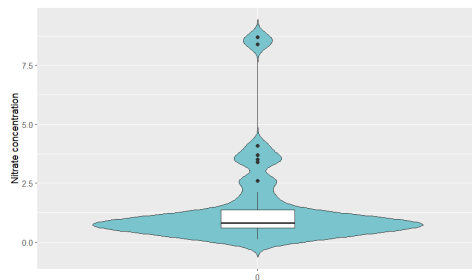


Figura 7: Violinplot della variabile
Nitrate

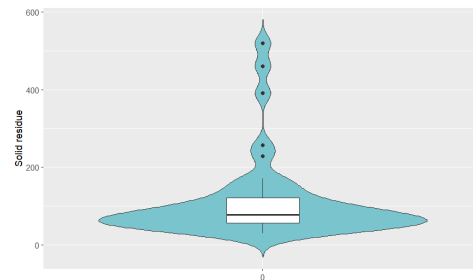


Figura 8: Violinplot della variabile
solid residue

La variabile **Nitrate** ha una distribuzione asimmetrica a destra (*Figura 7*), con un valore minimo di 0.100 ppm (*parts per million*) e un massimo di 8.700 ppm. La mediana è di 0.800 ppm, la media è di 1.495 ppm. la variabile **Solid Residue** ha anche essa una distribuzione asimmetrica a destra (*Figura 8*), con un valore minimo di 29.0 ppm e un massimo di 520.0 ppm. La mediana è di 78.0 ppm, la media è di 112.4 ppm (*Tabella 1*).

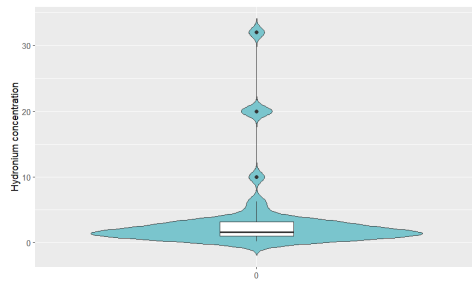


Figura 9: Violinplot della variabile
hydronium

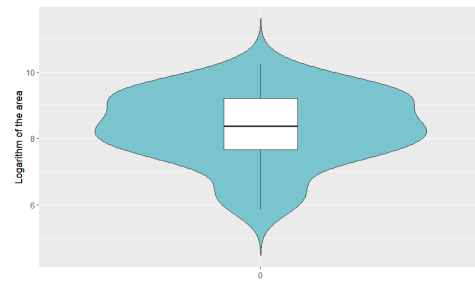


Figura 10: Violinplot della variabile
 $\ln(\text{Area})$

La variabile **Hydronium** ha una distribuzione asimmetrica a destra (*Figura 9*), con un valore minimo di 0.200 e un massimo di 32.00. La mediana è di 1.600, la media è di 3.631. La variabile **$\ln(\text{Area})$** segue una distribuzione simmetrica (*Figura 10*), con un valore minimo di 5.855 e un massimo di 10.24. La mediana è di 8.370, la media è di 8.331 (*Tabella 1*).

1.2 Analisi Bivariate

Si valuta la relazione tra le varie variabili. Delle 35 bivariate totali, 17 bivariate sono risultate significative (*Tabella 2*).